Development of omics-based tests for clinical use: the challenge of achieving statistical robustness and clinical utility

FDA Proteomics in the Clinic Workshop

Lisa McShane, PhD
Biometric Research Branch, DCTD
National Cancer Institute

June 13, 2014

Disclosures

- I have no financial relationships to disclose.
- I will not discuss off label use and/or investigational use in my presentation.
- The views expressed represent my own and do not necessarily represent views or policies of the National Cancer Institute.

My perspective

- Statistical/scientific reviewer of NCI-sponsored studies for development and validation of biomarker-based tests
- Scientific Advisory Board (*Science Translational Medicine*) and Editorial Board (*BMC Medicine*)
- Statistical collaborator in research projects

OUTLINE

- Background & definitions
- Roles for omics-based tests
- Define prognostic and predictive
- Two cases studies
  - Gene expression-based prognostic classifier in early stage lung cancer
  - Serum proteomic predictive classifier in advanced lung cancer
- Recommended reading

Working definitions

- **Biomarker**

  (http://www.cancer.gov/dictionary):
  "Biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease."

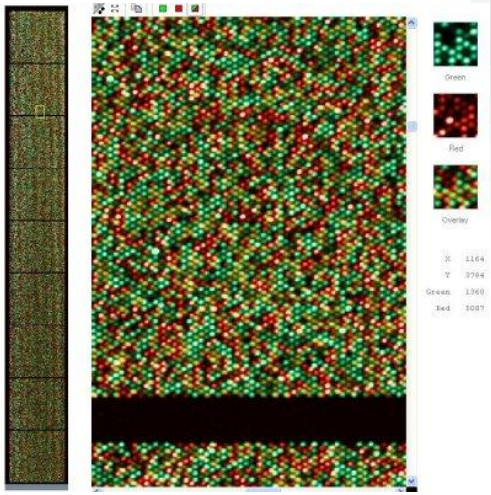
- **Omics**

  (http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx)

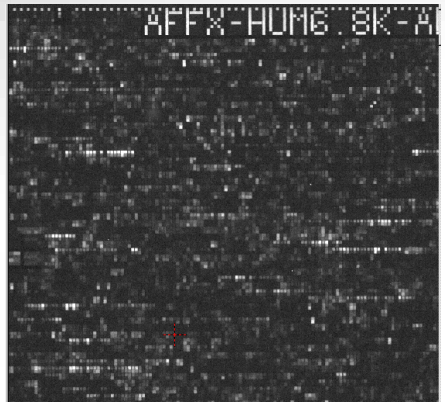
  "A term encompassing multiple molecular disciplines, which involve the characterization of global sets of biological molecules such as DNAs, RNAs, proteins, and metabolites."

4

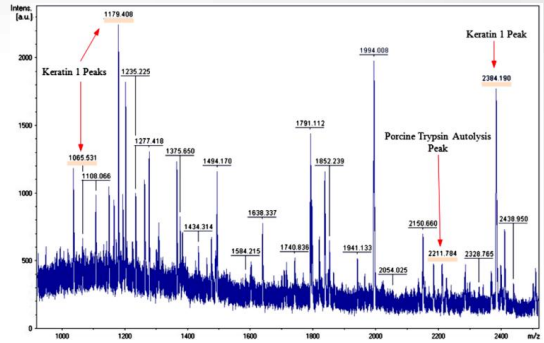
Many examples of biomarkers/omics for characterization of biological samples



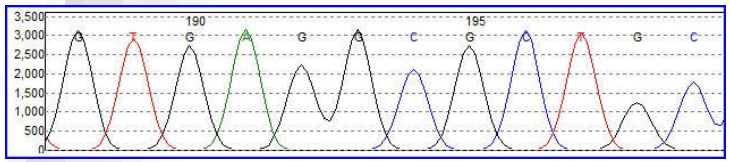
Illumina SNP bead array



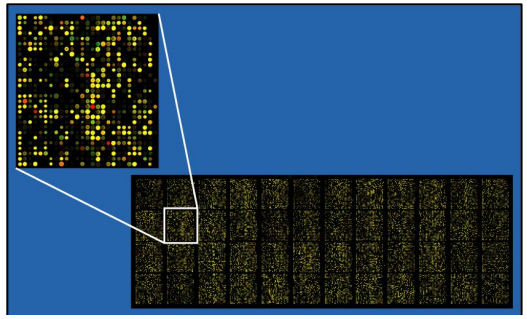
Affymetrix expression GeneChip



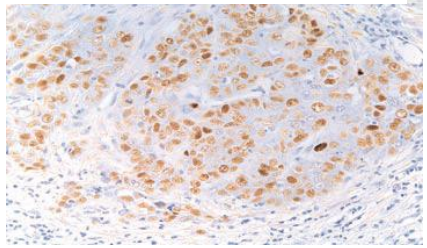
MALDI-TOF proteomic spectrum



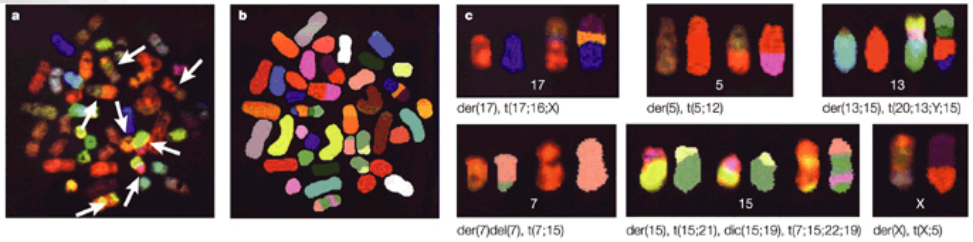
Mutation sequence surveyor trace



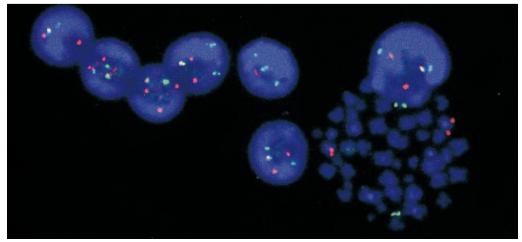
cDNA expression microarray



p53 IHC stain of breast cancer

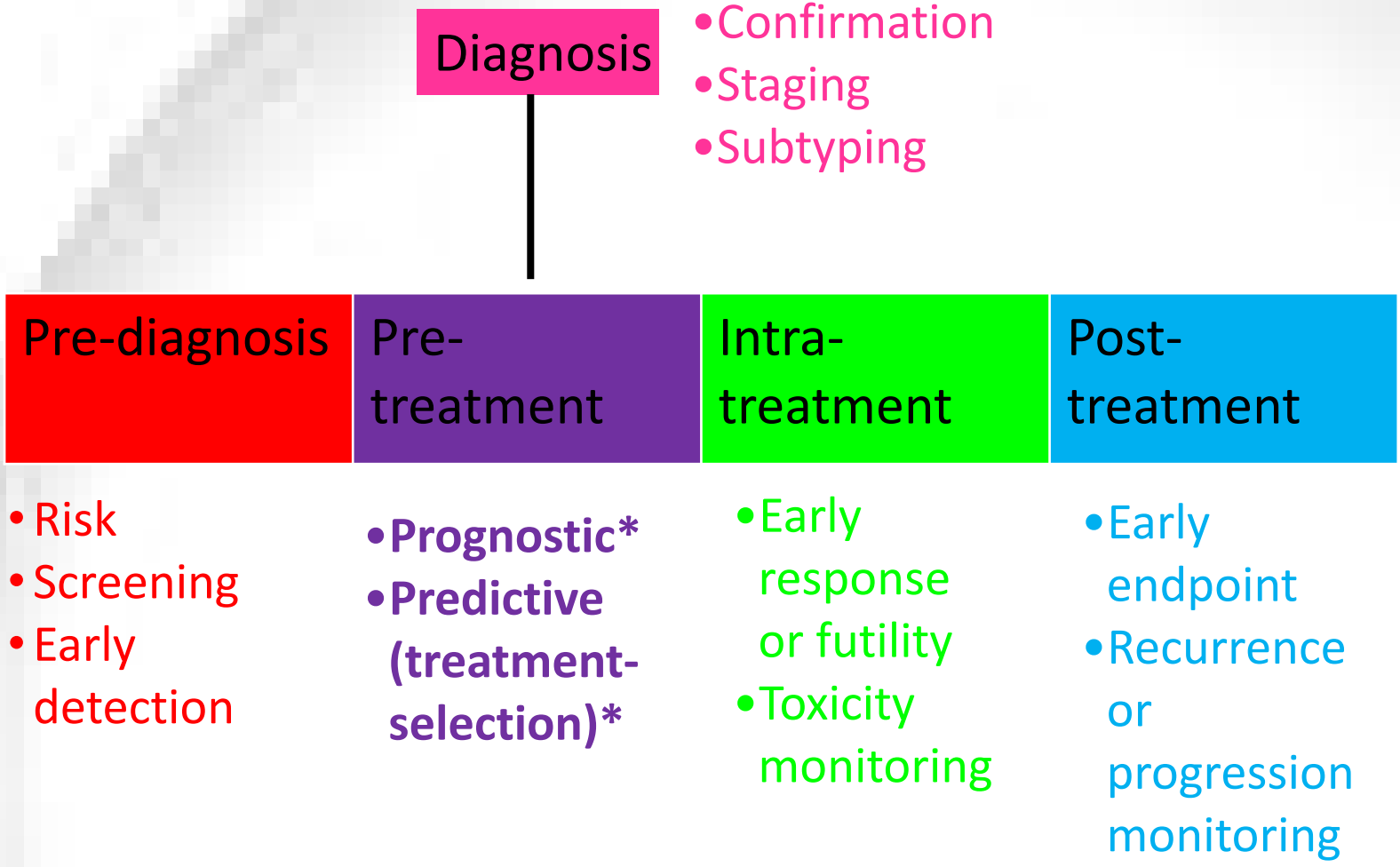


SKY analysis of AML cells



FISH analysis of BCR-ABL in ALL

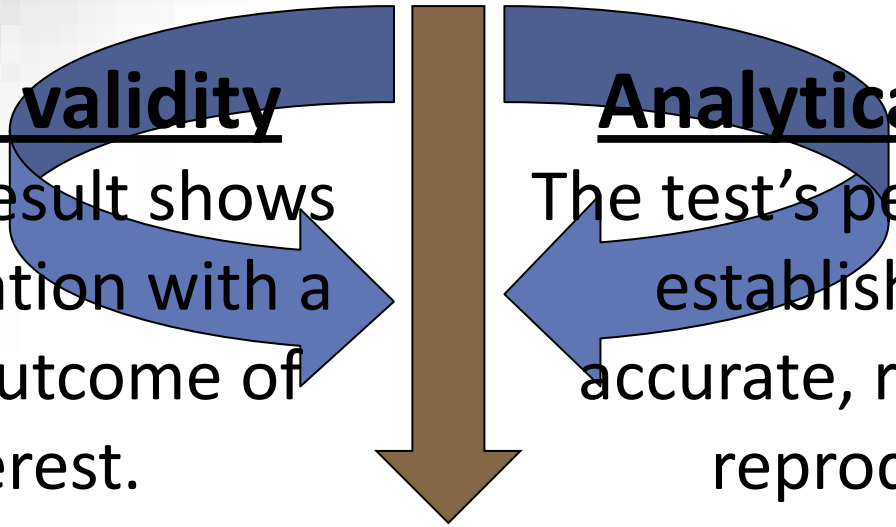
Potential roles for omics/biomarker-based tests



***Examples in this talk focus on tests for initial therapy selection.**

Paradigm for development of a clinically useful biomarker-based test

Discovery



Clinical validity

The test result shows an association with a clinical outcome of interest.

Analytical validity

The test's performance is established to be accurate, reliable, and reproducible.

Clinical utility

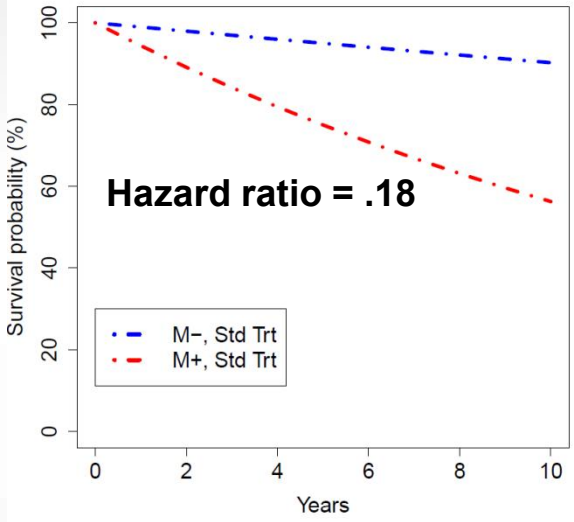
Use of the test results in a favorable benefit to risk ratio for the patient

Genet Med 2009;11:3-14
J Clin Oncol 2012;30:4223-4232

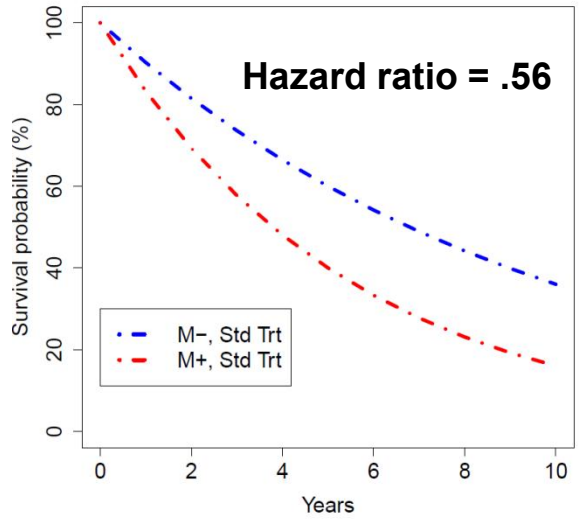
Prognostic biomarker

- Associated with clinical outcome in absence of therapy (natural course) *or with standard therapy all patients are likely to receive*
- Not always relevant for therapy decisions

Good prognosis group (M-) may forego additional therapy



Is this prognostic information helpful ?



M=biomarker

8

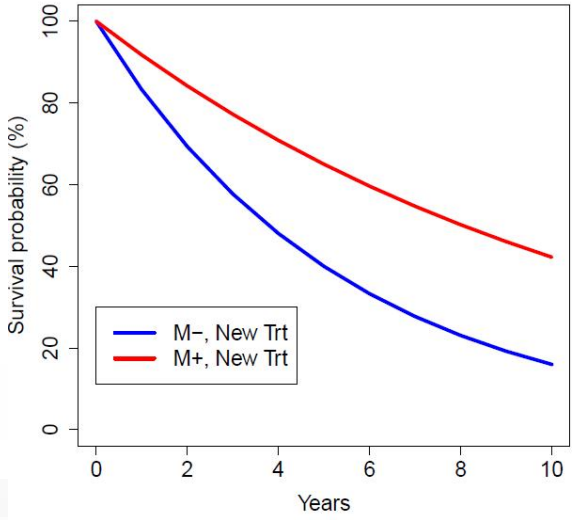
Predictive biomarker

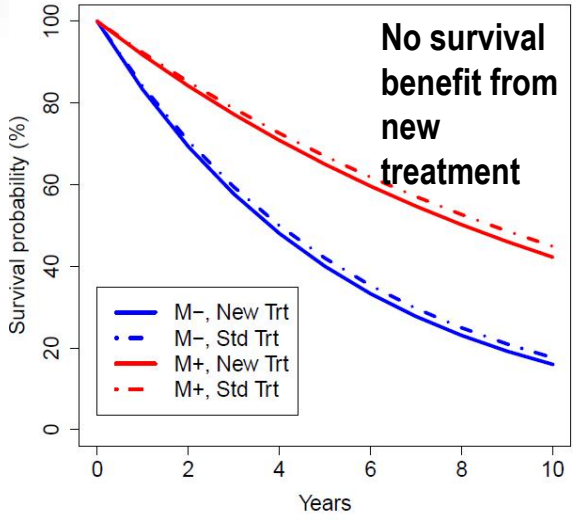
- Associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy
  - Alternate terms: treatment-selection, treatment-guiding, treatment effect modifier
- Generally more useful than prognostic biomarkers for therapeutic decision making

J Natl Cancer Inst 2013;105:1677-1683
Clinical Trials 2013; 10: 653-665

9

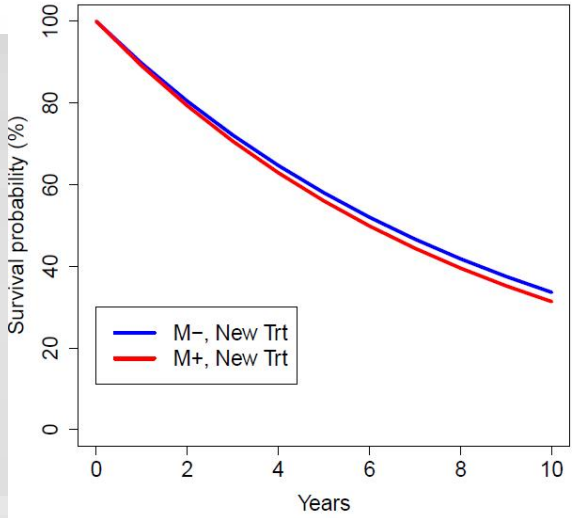
Prognostic vs. predictive: Importance of control groups

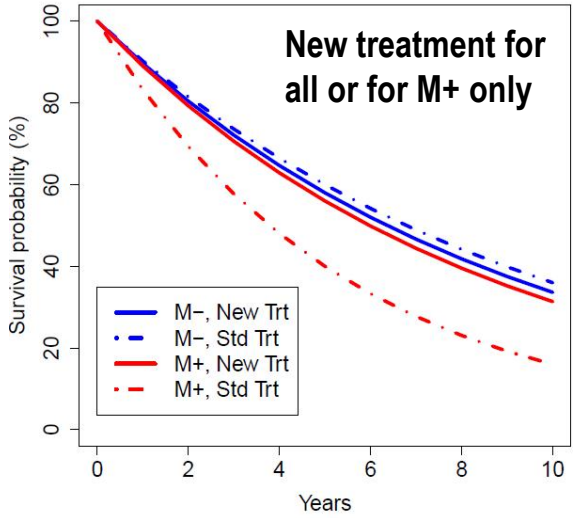




Prognostic but not predictive

(M = biomarker)





Prognostic and predictive

Statistical language for predictive biomarkers: "Treatment-by-biomarker interaction"

- Treatment effect (e.g., hazard ratio) varies by biomarker status
  - **Quantitative** interaction: Treatment benefits all patients but by different amounts
  - **Qualitative** interaction: Patients "positive" for the biomarker benefit from the treatment but others receive no benefit or possibly even harm

11

Plasma IL-6 as predictive biomarker for pazopanib vs. placebo?

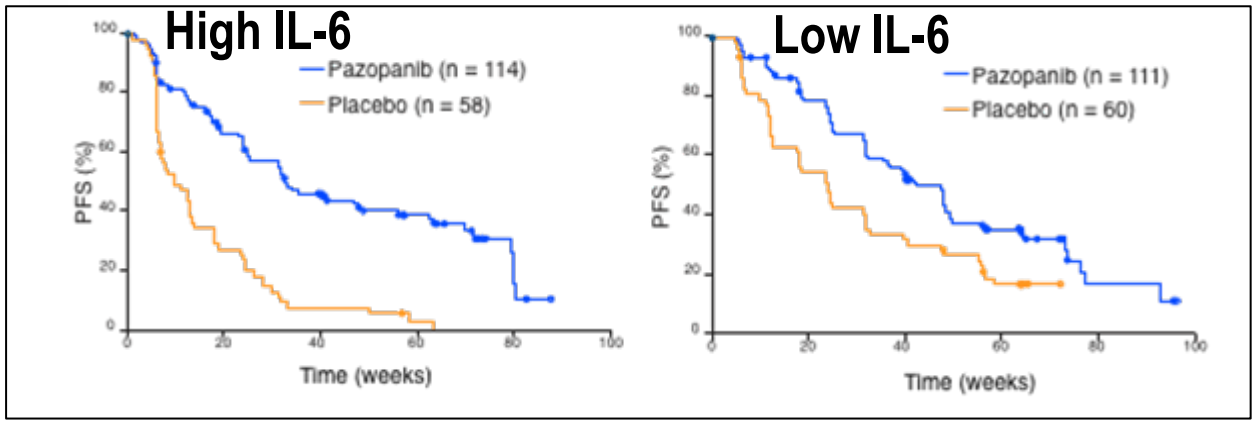
Results of randomized placebo-controlled phase III trial in metastatic renal-cell cancer

	PFS (weeks)		HR (95% CI)	p value		
	Pazopanib	Placebo		Pazopanib	Placebo	Interaction
Interleukin 6						
Low	42·3	24·0	0·55 (0·38–0·81)	0·445	<0·0001	0·009
High	32·6	9·9	0·31 (0·21–0·44)			

Prognostic: P<0.0001

Predictive?

Quantitative interaction: P=0.009



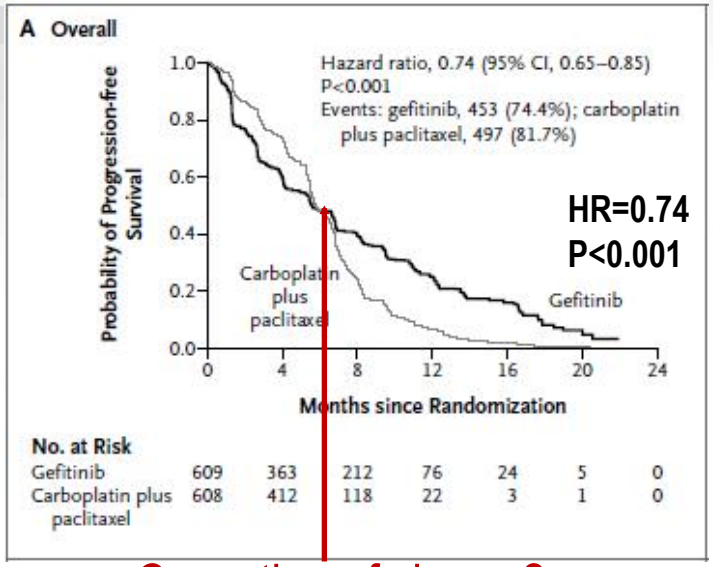
Lancet Oncol 2012;13:827-837

12

EGFR mutation predictive for PFS benefit with gefitinib in NSCLC

EGFR mutation is:

- Prognostic (positive)
- Predictive: **Qualitative** interaction, p<0.001)

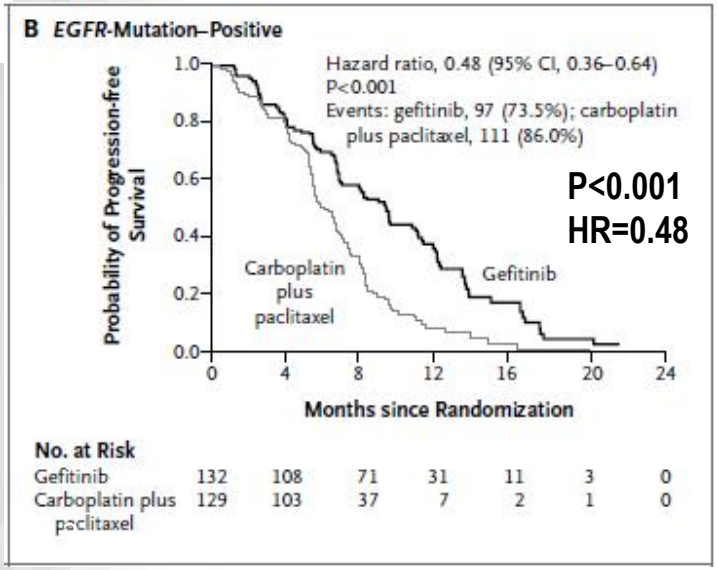


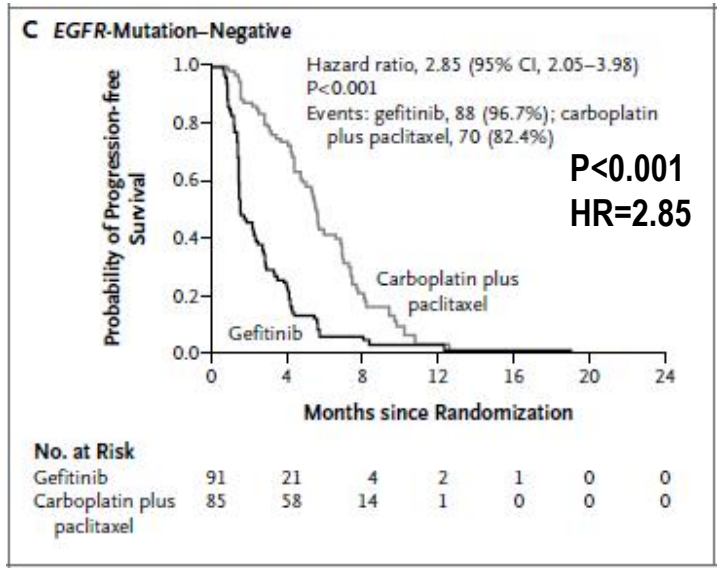
Cessation of chemo?

IPASS: Phase III 1st line advanced adeno NSCLC

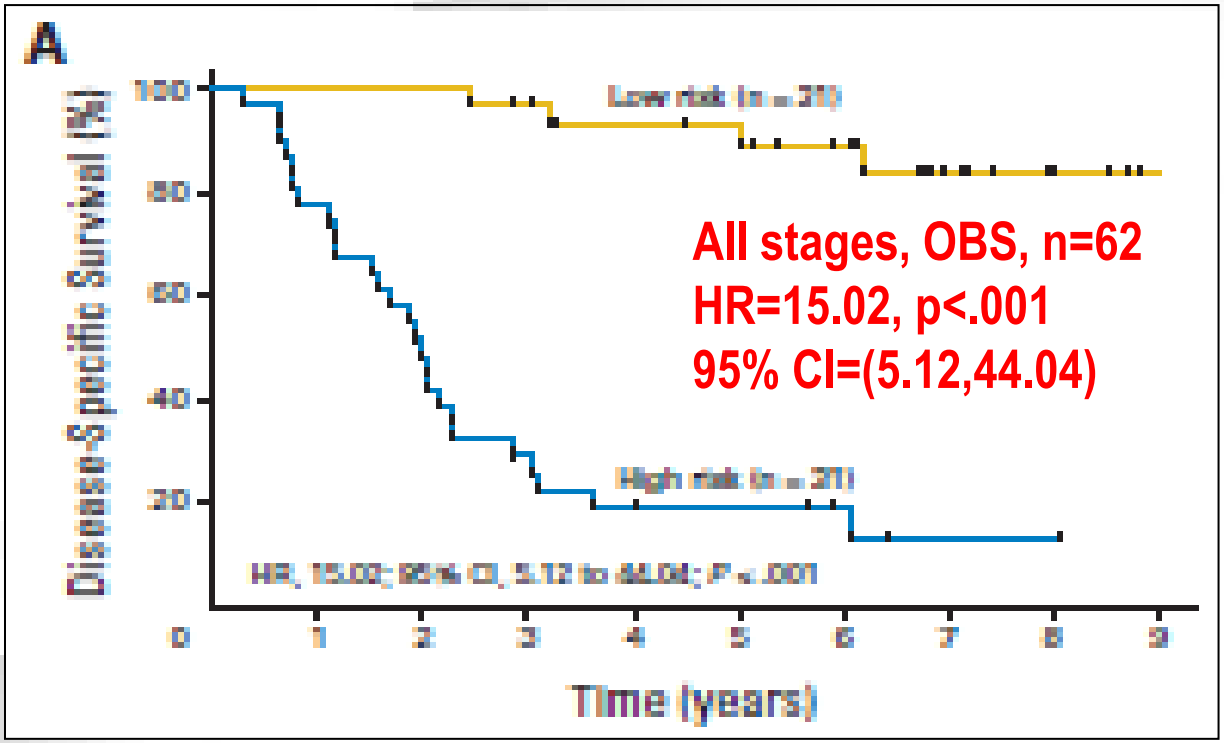
gefitinib vs. carboplatin+paclitaxel

(N Engl J Med 2009;361:947-57)





Prognostic classifier for early stage non-small cell lung cancer



A 15-gene signature was constructed using data from **OBS** arm of a randomized clinical trial (OBS vs. ACT) for lung cancer patients who were candidates for adjuvant chemotherapy.

"A 15-gene signature separated **OBS** patients into high-risk and low-risk subgroups with significantly different survival (hazard ratio [HR], 15.02; 95% CI, 5.12 to 44.04; *P <.001.*"
(J Clin Oncol 2010; 28: 4417-4424*)*

RESUBSTITUTION

Model development

Overfitting models built from high-dimensional (e.g., "omics") data

- A statistical model is **OVERFIT** when it describes random error or noise instead of the true underlying relationship
  - Excessively complex (too many parameters or predictor variables)
  - Generally has poor predictive performance on an independent data set
- **RESUBSTITUTION** is the naïve practice of evaluating performance of a model by "plugging in" exact same data used to build it

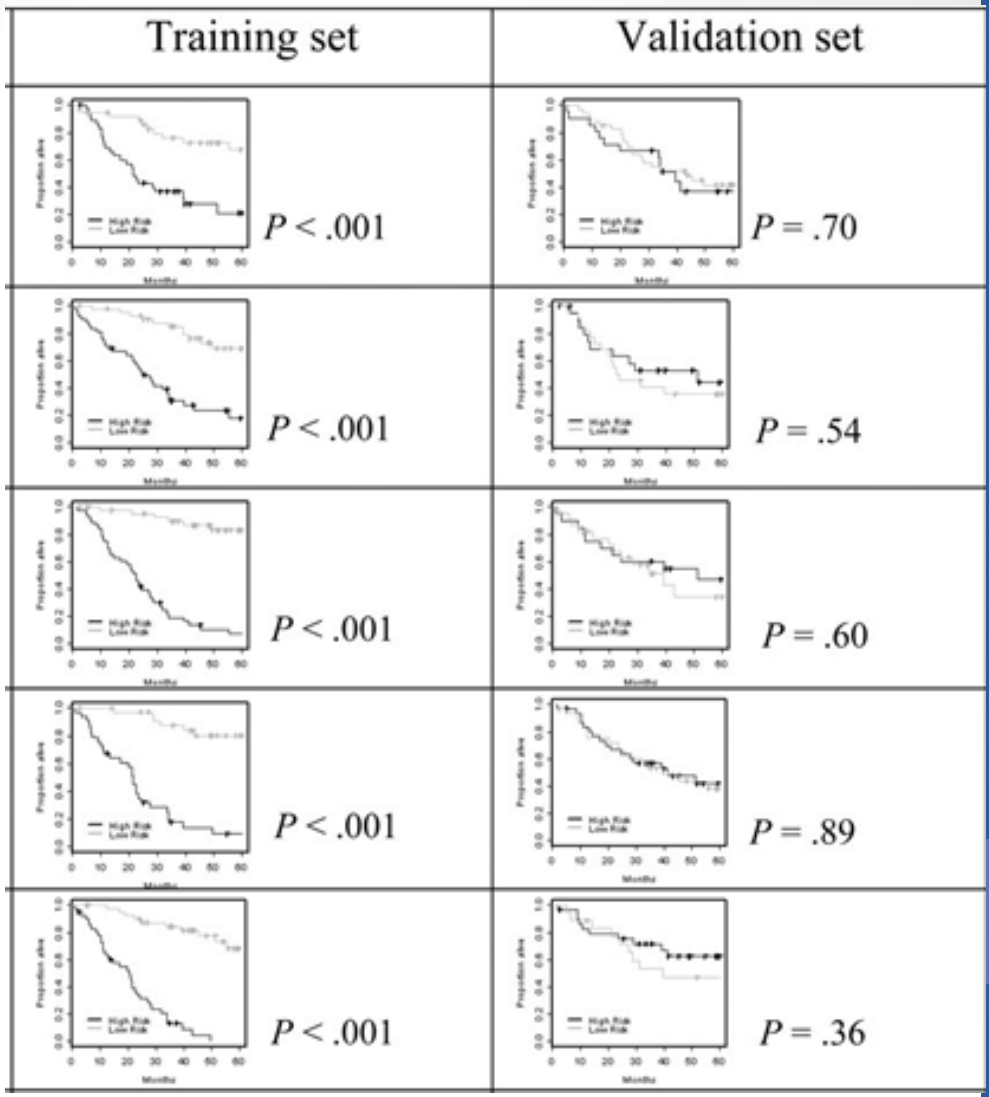
15

Model development

Model "resubstitution" pitfall

(Explained in J Natl Cancer Inst 2010; 102:464-474)

- Goal: Develop prognostic signature from gene expression microarray data
- Survival data on 129 lung cancer patients (prior study)
- Expression values for 5000 genes generated randomly from N(0, I_{5000}) ("noise") for each patient
- Data divided randomly into training and validation sets
- Prognostic model developed from training set and used to classify patients in both training and validation sets (supervised principal components method)

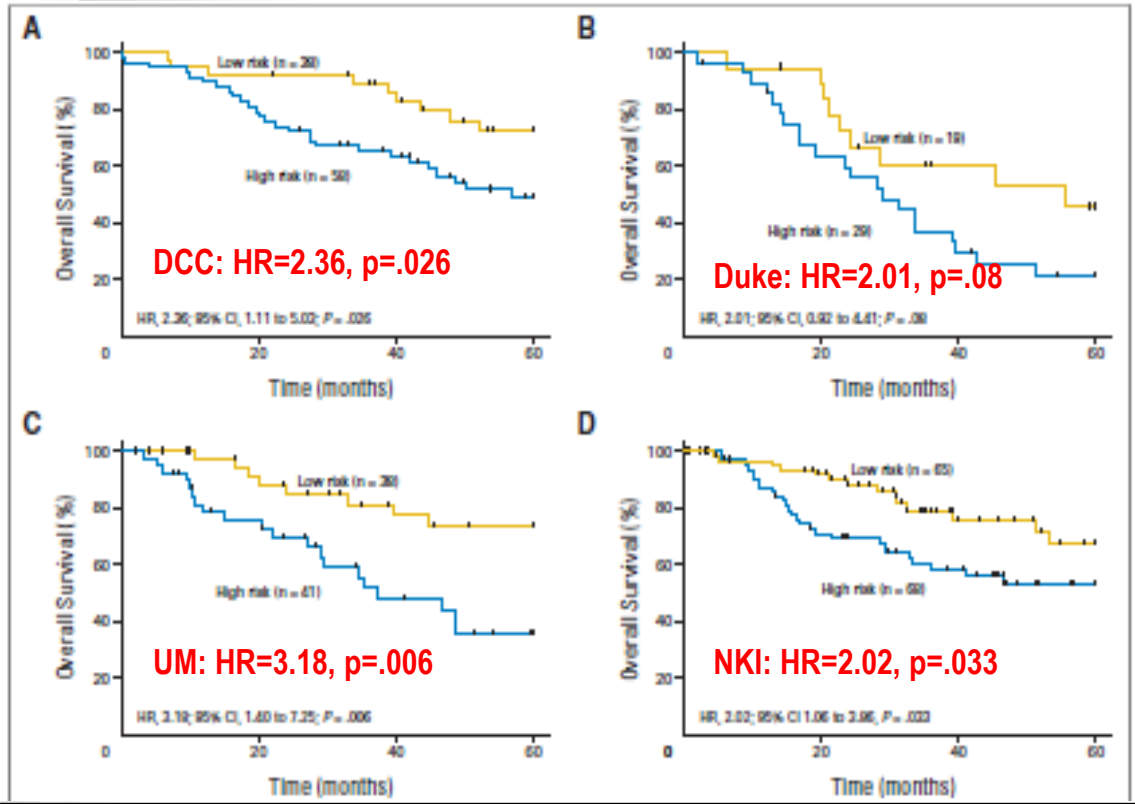


16

Prognostic classifier for early stage non-small cell lung cancer

Did it really validate?

" . . . prognostic effect was validated consistently in four separate microarray data sets (total 356 stage IB to II patients without adjuvant treatment)."



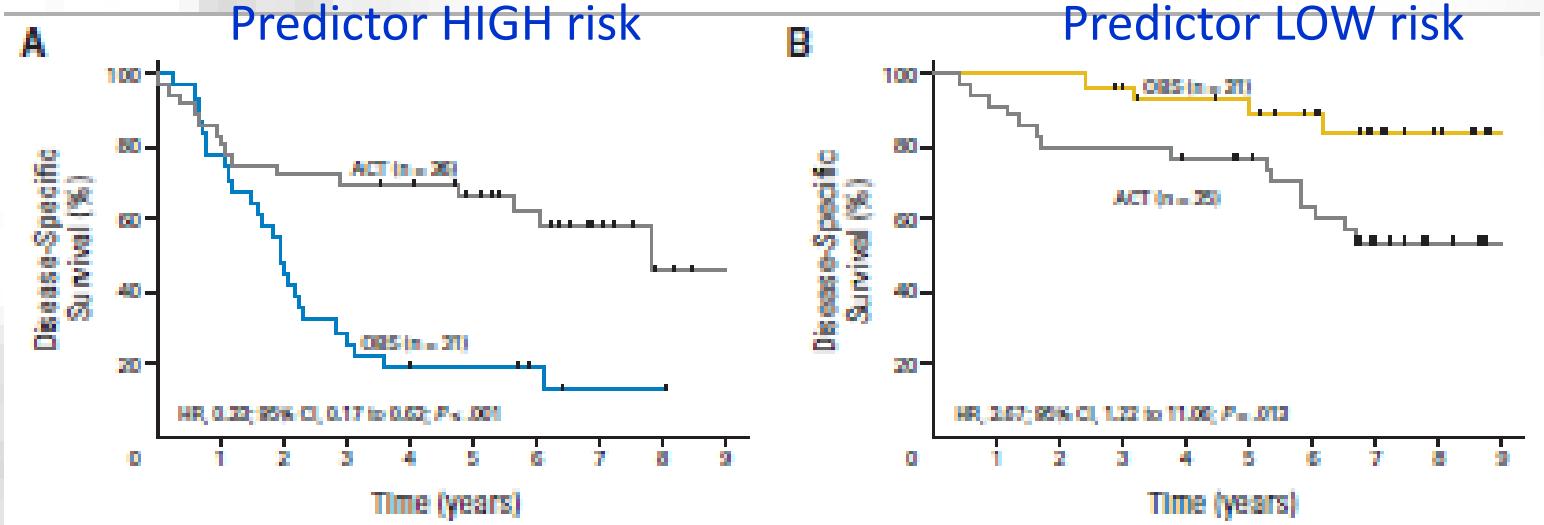
- What happened to HR=15.02?
- Endpoint: DSS→OS
- Timescale: 9 →5 yrs
- Mixed stages

Prognostic classifier for early stage non-small cell lung cancer

Is it also predictive?

Predictor HIGH risk

Predictor LOW risk



"The signature was also predictive of improved survival after ACT in JBR.10 high-risk patients (HR, 0.33; 95% CI, 0.17 to 0.63; P =.0005), but not in low-risk patients (HR, 3.67; 95% CI, 1.22 to 11.06; P = .0133; interaction P < .001)." (J Clin Oncol 2010; 28: 4417-4424)

RESUBSTITUTION strikes again

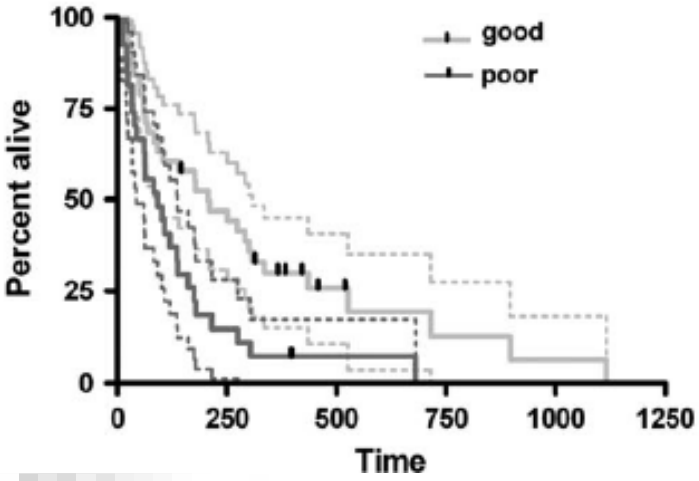
Model development: Serum proteomic test to classify NSCLC for outcome with EGFR-TKIs

- Serum collected from NSCLC patients before treatment with gefitinib or erlotinib (EGFR-TKIs)
- Analysis by MALDI-MS
- K-nearest neighbor (KNN) algorithm based on 8 distinct m/z features classifies into good or poor outcome
- Training set: n=139 NSCLC patients total from 3 cohorts who received gefitinib
- Preliminary validation cohorts:
  - "Italian B": n=67 sequential patients, late-stage or recurrent NSCLC treated with single-agent gefitinib
  - ECOG 3503: n=96 advanced NSCLC patients treated with first-line erlotinib on single arm Phase II study

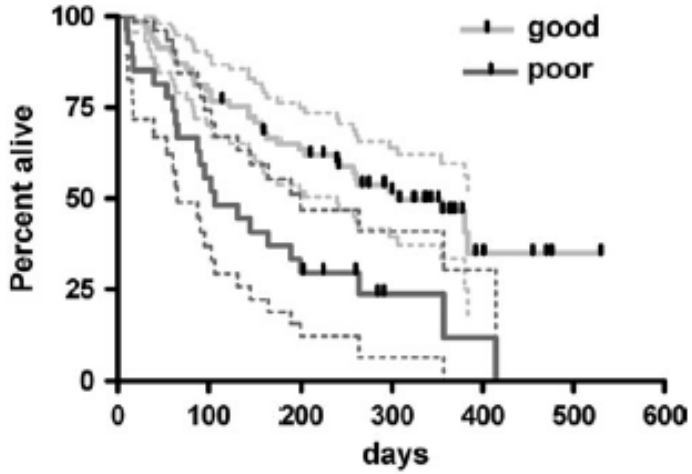
J Natl Cancer Inst 2007;99:838-46

Preliminary validation: Proteomic test to classify NSCLC for outcome with EGFR-TKIs

Preliminary results for patients treated with EGFR-TKIs



"Italian B": n=67 sequential patients, late-stage or recurrent NSCLC treated with single-agent gefitinib
HR=0.50, 95% CI=(0.24,0.78), p=0.0054
Median OS
Good: 207 days Poor: 92 days

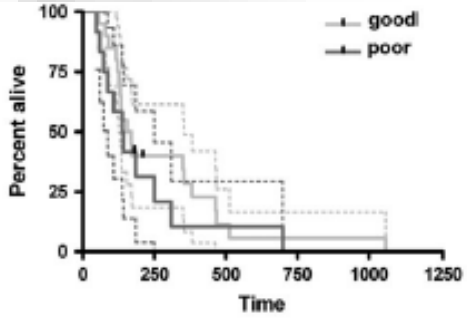


ECOG 3503: n=96 advanced NSCLC patients treated with first-line erlotinib on single arm Phase II study
HR=0.4, 95% CI=(0.24,0.70), p<0.001
Median OS
Good: 306 days Poor: 107 days

Proteomic test shown to have good analytical reproducibility across 2 labs

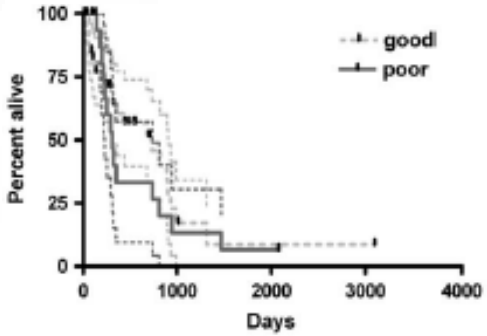
Predictive or Prognostic? Proteomic test to classify NSCLC for outcome with EGFR-TKIs

Does test also separate by outcome patients who did NOT receive EGFR-TKIs (control cohorts)?



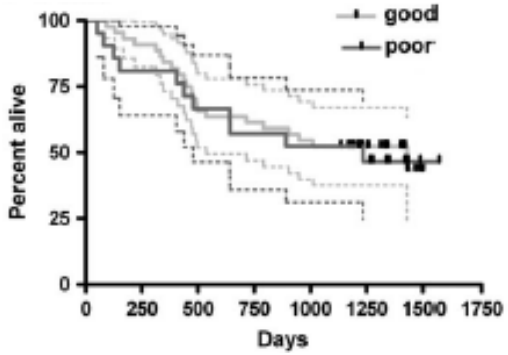
"Italian C": n=32 patients, stage IIIA-IV NSCLC treated with second-line chemotherapy
HR=0.74, 95% CI=(0.33,1.6), p=0.42

SAME TREND, BUT NS



"VU": n=61 patients, advanced NSCLC treated with second-line chemotherapy
HR=0.81, 95% CI=(0.4,1.6), p=0.54

SAME TREND, BUT NS



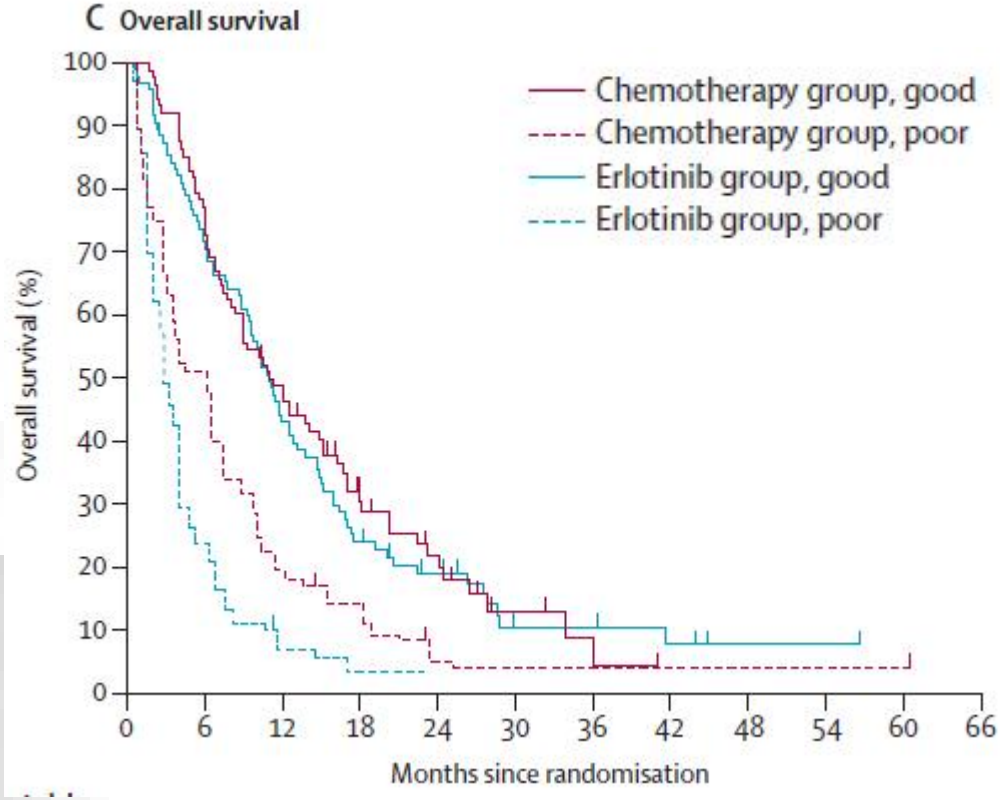
"Polish": n=65 patients, stage IA-IIB NSCLC treated with second-line chemotherapy
HR=0.90, 95% CI=(0.43,1.89), p=0.79

SAME TREND, BUT NS

Randomized phase III trial (PROSE): Proteomic test to classify NSCLC for outcome with EGFR-TKIs

- Test predictive value of the proteomic test
- Primary endpoint overall survival (OS)
- Powered for treatment x proteomic test interaction
- Eligibility
  - Stage IIIB or IV NSCLC
  - ≥ 18 years old
  - Refractory to one prevision platinum-containing regimen
- Exclusions
  - Previously received an EGFR-TKI
  - Uncontrolled brain metastases
  - Other cardiac, renal, etc. conditions

Randomized phase III trial (PROSE): Proteomic test to classify NSCLC for outcome with EGFR-TKIs



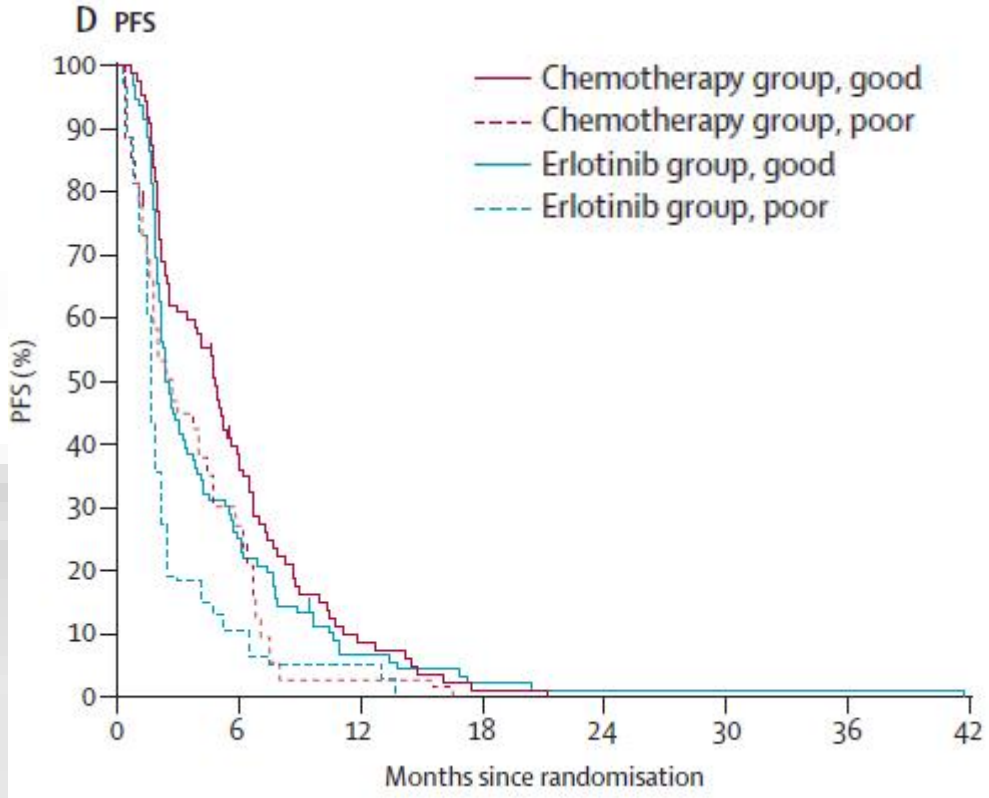
Median Overall Survival (months)

	Test result	
Treatment	Good	Poor
Chemo	10.9	6.4
Erlotinib	11.0	3.0
Hazard ratio (95% CI)	1.06 (0.77-1.46)	1.72 (1.08-2.74)

Interaction p=0.017

"Serum protein test status is predictive of differential benefit in overall survival for erlotinib versus chemotherapy in the second-line setting. Patients classified as likely to have a poor outcome have better outcomes on chemotherapy than on erlotinib." (Lancet Oncol 2014;15:713-21)

Randomized phase III trial (PROSE): Proteomic test to classify NSCLC for outcome with EGFR-TKIs



Median Progression-Free Survival (months)

	Test result	
Treatment	Good	Poor
Chemo	4.8	2.8
Erlotinib	2.5	1.7
Hazard ratio (95% CI)	1.26 (0.94-1.96)	1.51 (0.96-2.38)

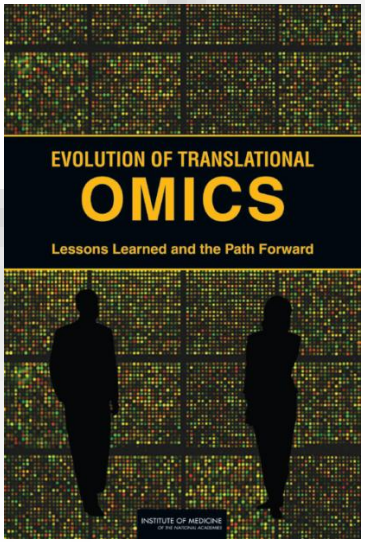
Interaction p=0.445

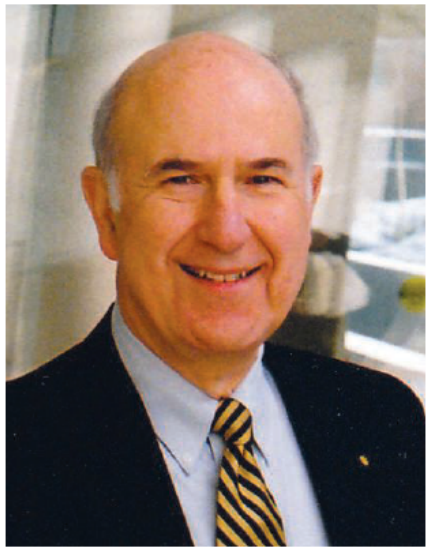
The indication for the test seems to have drifted from a test to select who will benefit from erlotinib to who should receive chemotherapy.

Proteomic test to classify NSCLC for outcome with EGFR-TKIs : Many questions remain

- Impact of patient selection criteria for trial (patients could not have prior EGFR-TKI)
- Impact of subsequent therapies on OS endpoint
- Important differences in drug delivery (oral vs. IV)
- Important differences in toxicity profile
- Is giving all patients chemotherapy a reasonable option?

Institute of Medicine report on the field of translational omics





"There are a lot of lessons here that surely apply to other places."

—GILBERT S. OMENN, UNIVERSITY OF MICHIGAN, ANN ARBOR

http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx

NCI criteria for the use of omics-based predictors in clinical trials.
Nature 502: 317-320, 2013.
BMC Medicine 11:220, 2013.

Thanks for your attention!